Structuring Healthcare Knowledge Bases:
An Analysis of Explicit and Implicit Structures in Arden Syntax and An XML Schema Representation of Arden Syntax

Ashwin Jadhav and Matthew Sailors
School of Health Information Sciences and Medical School Department of Surgery
The University of Texas Health Science Center at Houston

The Arden Syntax for Medical Logic Modules (MLMs) is an ANSI and ISO recognized standard language for representing clinical knowledge bases. We analyzed the explicit and implicit structures in Arden Syntax MLMs and developed an information model represented as an XML schema. While the explicit structures were easily represented as XML, implicit structures require further explicit definition. In any future representation format, explicit structuring must be balanced with expressiveness and usability.

METHODS

An analysis of the category and slot structures of an Arden Syntax Medical Logic Module[1,2] was completed and a preliminary information model based on the explicit and implicit structures as defined by the textual portions of the Arden Syntax standard was developed. Each Arden MLM has three categories: maintenance, library, and knowledge. The categories are further divided into slots. The slots were analyzed for data-type, pattern, and syntax definition and also for utility of further explicit syntax definition. This model was refined to include structured information within the maintenance and library category slots. The model was further refined after reviewing the Arden Syntax context-free grammar. The refined model was converted into an XML Schema[3] for further review and refinement.

STRUCTURES

The entire MLM has explicit constraints on the structure of the categories and slots. Additional constraints limit the recommended character set.

The maintenance category slots (title, mlmname, arden, version, institution, author, specialist, date, and validation) all have explicit data type and pattern definitions (allowed values; constraints). Title, mlmname, version, and institution also have implicit syntax definition.

Within the library category, the purpose and explanation slots cannot be rigidly structured since they have free text content. However, the keywords, citations, and links slots could be further structured to facilitate processing and use of the information in these slots.

The knowledge category slots type, data, priority, evoke, logic, action, and urgency all have explicitly defined data-types, patterns, and syntaxes. The complex nature of the data, logic, and action slots make explicit structuring of the statements in these slots challenging. The modeling of these statements was outside of the scope of this project but will be explored during future work. Further work to structure these slots will need to balance completeness and expressiveness against usability.

XML SCHEMA

An XML Schema based on our analysis of structures of Arden Syntax was developed. This schema models the explicit and implicit structures as defined by the Arden Syntax standard and extends the implicit structuring using other models suggested by the Arden Syntax standard.

This XML model could be used to increase the portability of Arden MLMs by simplifying the conversion into other document types. We found that explicit structures in Arden Syntax are easy to model in a Schema, although the implicit structures may require more detailed definitions to be truly usable.

The resulting schema was presented to the Arden Syntax Special Interest Group of Health Level Seven, Inc. at the January 2003 Working Group Meeting for consideration as part of the next revision of the Arden Syntax Standard. A revised schema was presented in May 2003 and will be incorporated in the new draft of "Arden 2.x".

References

1. Sailors RM, Jenders RA. Arden Syntax 2.1: Ten Years of Standardized Knowledge Bases. AMIA Annual Fall Symposium 2002.
2. Health Level Seven, Inc. Arden Syntax for Medical Logic Systems Version 2.1. Health Level Seven, Inc., Ann Arbor, MI, 2002.
3. World Wide Web Consortium Recommendations Extensible Markup Language (XML) 1.0 (Second Edition).